

# Investigación reproducible: Promoción y uso en educación superior

## Reproducible research: promotion and use in higher education

Jesús Humberto Cuevas Acosta\*

\*Doctor en Educación. Profesor Investigador del Tecnológico Nacional de México, Especialización en Docencia. Maestro en Ciencias en Enseñanza de la Ciencia con especialidad en Física. Ingeniero Industrial, [jesus.humberto.cuevas@outlook.com](mailto:jesus.humberto.cuevas@outlook.com)

Cómo citar / How to cite

Cuevas, J. (2017). Investigación Reproducible: Promoción y Uso en Educación Superior. *Yulök Revista de Innovación Académica*, 1(1), 9-17.

### Resumen

Este artículo presenta los resultados de un estudio que exploró en que medida se promueve y utiliza la reproducibilidad de la investigación entre estudiantes de Ingeniería. Se partió de dos premisas: [1] la institución escolar es un entorno favorable para promover la reproducibilidad, y [2] el profesor es un factor esencial en el fomento de la honestidad intelectual. Se diseñó y aplicó un cuestionario de 14 ítems a 600 estudiantes. Los resultados encontrados indican que la promoción de la reproducibilidad es muy baja; una cantidad representativa de los estudiantes ignoran la existencia de lenguajes de marcado como LATEX– Markdown y de cómputo estadístico como R. Se recomienda elaborar y someter a prueba modelos de trabajo para introducir la reproducibilidad como parte inherente de la formación de profesionales en el campo de la Ingeniería.

**Palabras clave:** investigación reproducible, educación superior, lenguajes de programación

### Abstract

This paper presents the results of a study that explored to what extent it is promoted and used the reproducibility of research among engineering students. It started from two premises: [1] The school institution is a favorable environment to promote reproducibility, and [2] The teacher is an essential factor in the promotion of intellectual honesty. A questionnaire of 14 items was designed and applied to 600 students. The results indicate that the promotion of reproducibility is very low; a representative quantity of students don't have knowledge of the existence of markup languages like LATEX– Markdown and statistical computing language like R. It's recommended the develop and test working models for introduction the reproducibility as an inherent part of the training of professionals in the field of engineering.

**Keywords:** reproducible research, higher education, programming languages

### Introducción

Cuando se realiza un estudio que generó conocimiento de frontera, un dispositivo tecnológico, aplicaciones innovadoras o nuevos métodos y técnicas de operación, el investigador o equipo de trabajo documentan sus resultados y planean de manera rigurosa su difusión.

Habitualmente, el protocolo de trabajo que siguen se asemeja al siguiente: [1] Examinan los medios más adecuados para difundir sus hallazgos, privilegiando aquellos que permitan mayor visibilidad, alta especialización e impacto internacional. [2] Elaboran un reporte escrito – frecuentemente un artículo– atendiendo lineamientos establecidos por una comunidad epistémica de reconocido prestigio. [3] Someten el reporte a evaluación experta por



pares independientes, quienes analizan la congruencia del método seguido respecto de las interrogantes, objetivos e hipótesis planteadas; revisan los resultados obtenidos y comparan la interpretación y coherencia de los mismos con la discusión realizada; juzgan la pertinencia de las fuentes de información utilizadas y valoran de forma integral el aporte a la disciplina científica en el que se enmarca el estudio. [4] De forma paralela se efectúa una declaración jurada en que él o los autores manifiestan que sus resultados son originales, no se han publicado en otros medios ni se encuentran en proceso de valoración alguna. [5] Si los resultados de la evaluación son positivos, el artículo se difundirá entre la comunidad científica que, a su vez, podrá evaluar, comentar e incluso reproducir los resultados siguiendo el método descrito para validar o refutar el estudio en su conjunto.

Según Gandrud (2015), los reportes de investigación se presentan de formas muy selectivas como artículos en revistas especializadas, libros e incluso sitios web. Otras maneras de presentación son conferencias, paneles de expertos y carteles en foros académico—científicos.

No obstante, durante las últimas dos décadas se han incrementado las denuncias relacionadas con la veracidad de los resultados de investigación en distintas áreas de la ciencia y el desarrollo tecnológico, especialmente las que implican el uso de actividades experimentales. En muchas disciplinas científicas se han detectado dificultades para reproducir los resultados publicados, aunque destacan los efectuados en las áreas de Biología, Medicina y Psicología.

En un estudio sobre reportes en el campo de la Psicología, la Open Science Collaboration (OSC) publicó un artículo con los resultados de un escrutinio de 100 reportes de investigación publicados en revistas de alto impacto; encontró que 61 no pudieron reproducirse y en el resto los resultados no fueron concluyentes (Open Science Collaboration, 2015).

También Iqbal (2016) y un equipo de colaboradores señalaron la falta de transparencia en muchos reportes publicados. Uno de sus últimos trabajos consistió en analizar 441 estudios publicados en el periodo 2000–2014 y hallaron que únicamente cuatro tenían el propósito de replicar estudios previos y ningún autor o grupo de trabajo proporcionó datos que permitieran ser analizados por evaluadores independientes. Una década antes, Ioannidis (2005) había advertido a la comunidad científica internacional que la mayoría de los resultados que se publican en revistas científicas son falsos.

En un artículo publicado en la revista *Nature*, Baker (2016) entrevistó a 1,576 investigadores respecto a la reproducibilidad de la investigación. Encontró que más del 70% de los encuestados señaló dificultades para reproducir experimentos y resultados publicados por colegas de otros grupos de trabajo, que el 50% no pudo replicar los resultados en sus laboratorios y únicamente el 20% decidió contactar a los autores para informarles.

Los resultados de los reportes reseñados permiten inferir la crisis que atraviesa la ciencia y su difusión. También da un indicio de las dificultades del entorno en que deben efectuar su trabajo los investigadores, las exigencias cada vez mayores a las que están sometidos y por qué guardan silencio cuando no pueden reproducir resultados o replicar experimentos. Es probable que estas dificultades estén relacionadas con:

#### *Ausencia de apoyo para indagar los orígenes de un problema.*

Es difícil que una universidad, centro de investigación o laboratorio asignen recursos económicos y humanos suficientes para reproducir estudios con el objetivo de valorar la consistencia interna y usar los resultados como un punto de partida para investigaciones ulteriores.

#### *Limitación financiera.*

El enfoque pragmático predominante en las corporaciones e instituciones académicas exige realizar estudios que generen resultados tangibles, viables en términos económicos y factibles de generar utilidades en corto tiempo, limitando los recursos para procesos de reproducción.

#### *Inseguridad intelectual.*

Cuando el investigador no consigue reproducir los resultados, posiblemente se plantee las siguientes interrogantes: ¿estoy siguiendo el protocolo experimental de forma precisa? ¿existen discrepancias en los instrumentos de medición usados en el proceso de reproducción? ¿en qué medida la no reproducibilidad se debe a inconsistencias en el método descrito, a mi falta de pericia o ambas?

#### *Necesidad de publicar en las revistas de “alto impacto” científico.*

Por tradición, la cantidad de artículos publicados, la relevancia científica de las revistas y el número de “citas” recibidas, constituyen parámetros para medir el éxito de un investigador, la consolidación de un grupo de trabajo y el prestigio de una corporación o institución. En consecuencia, el investigador es presionado para publicar artí-

culos con resultados novedosos en medios de este tipo si es que desea acceder a fuentes de financiamiento, aspirar a tener un contrato laboral con mayor estabilidad económica o mantener su empleo actual.

Por otra parte, no existe un consenso internacional sobre la acepción de los términos reproducción y replicación. Habitualmente se reconoce que Jon Claerbout y Martin Karrenbach acuñaron el término investigación reproducible (Reproducible Research) en una conferencia dictada en el Meeting of The Society of Exploration Geophysics (Claerbout y Karrenbach, 1992) y por sus contribuciones posteriores en (Fomel y Claerbout, 2009).

¿En qué consiste la reproducción? ¿qué es la replicación? ¿cuáles son las diferencias y similitudes entre estos términos en el ámbito de la investigación? ¿en qué medida están interrelacionadas? son interrogantes que continúan generando un amplio debate entre la comunidad científica internacional.

En este artículo se asume que los términos tienen acepciones distintas y claramente diferenciadas que se acentúan en estudios de corte cuantitativo. Así, se proponen las definiciones siguientes:

*Reproducción* es repetir un estudio científico de forma independiente o no, respetando los objetivos, datos, métodos, protocolos experimentales y de análisis originales con el propósito de obtener los mismos resultados.

*Replicación* es conducir de manera independiente un proceso de recreación de hallazgos derivados de un estudio siguiendo el protocolo original de experimentación, acopiando y analizando datos nuevos para llegar a las mismas conclusiones.

Por tanto, reproducción y replicación se distinguen en que la primera se efectúa con procedimientos y datos originales y la segunda requiere que, además de respetar el protocolo de trabajo original, se obtengan datos nuevos.

Es importante advertir que al replicar una investigación se generan fuentes potenciales de error que deben mantenerse bajo control, entre ellas sobresalen: [1] cantidad, configuración y calibración del equipo en laboratorios; [2] insumos de trabajo con atributos distintos; [3] investigadores con habilidades, experiencia y formación científica diferente.

Un estudio reproducible no implica que sea replicable. La replicabilidad es más integral que la reproducibilidad. Por tanto, la capacidad de reproducir un estudio es el estándar mínimo que debe cumplirse cuando la replicación independiente no es factible o resulta difícil de efectuar

por cuestiones éticas, económicas, de espacio o tiempo. Tampoco implica que sea correcto; los errores también se reproducen y pueden pasar inadvertidos, por tanto, cuando se sospeche la existencia de fallos, es recomendable –de ser posible– utilizar datos nuevos e incluso modificar protocolos de trabajo.

Aún cuando la investigación reproducible tiene muchas ventajas en su implementación, existen limitaciones que deben tenerse en cuenta antes de iniciar un estudio que contemple la reproducibilidad total o parcial de sus resultados. Según Xie (2015), no se debe esperar que todos los reportes de investigación sean reproducibles ni que estén a disposición pública; sin embargo siempre será mejor compartir datos cuando sea posible.

Es difícil asegurar que proporción de reportes de investigación que no se han podido reproducir es por dominio disciplinar y metodológico insuficiente, protocolos diseñados de forma errónea, muestras no representativas, estadísticos de prueba inadecuados, o por la preferencia de las revistas especializadas en publicar estudios innovadores con resultados positivos en detrimento de resultados negativos o estudios anteriores que no han podido ser reproducidos. Más difícil aún es detectar si la imposibilidad de reproducir reportes se debe a conductas científicas inapropiadas relacionadas con creación de información falsa, alteración de datos, figuras e imágenes, manipulación de procedimientos experimentales, deducciones que los datos no permiten efectuar, conclusiones sobrevaloradas, entre otras. Las implicaciones de este problema son importantes si se asume que:

- La capacidad de reproducción, replicación y falsación son características inherentes del método científico.
- La honestidad intelectual es un atributo irreductible que debe promoverse socialmente sin distinción de ningún tipo.
- Un proceso formal y riguroso de reproducción exige al menos ser capaz de utilizar software especializado, métodos de análisis estadístico y organizar un reporte conforme a lineamientos metodológicos formales.

En consecuencia, cuando se elabora un protocolo de investigación que por sus características sea necesario estructurarlo para reproducirse, surgen las siguientes interrogantes: ¿En qué medida se conocen las herramientas computacionales necesarias? ¿En qué grado se cuenta con antecedentes en tópicos estadísticos y métodos de investigación?



El propósito de este estudio fue explorar en qué medida se promueve y utiliza la reproducibilidad de la investigación entre estudiantes de Ingeniería inscritos en instituciones universitarias. Se hizo énfasis en obtener información relacionada con el uso de software especializado, tópicos estadísticos y su articulación metodológica en la elaboración de reportes escolares de corte cuantitativo.

La comunidad estudiantil de Ingeniería constituyó el foco de atención por cinco razones fundamentales: [1] el tipo de problemas que deberán enfrentar en el ejercicio de su profesión; [2] la necesidad que tendrán de efectuar estudios observacionales y experimentales que exigirán el uso de métodos estadísticos en el análisis de los datos acopiados o generados; [3] el requisito de documentar resultados en formatos estandarizados y portables que no estén condicionados al uso de un sistema informático particular; [4] alta probabilidad de que la capacidad de reproducción o replicación de un estudio se convierta en un estándar internacional; [5] la exigencia de formar profesionales con un alto sentido de responsabilidad y honestidad que consideren la transparencia y colaboración como parte inherente de su actuar profesional y personal.

Para efectuar el estudio se asumió la tesis que considera la institución escolar como entorno fundamental para promover la reproducibilidad y al profesor como factor esencial en el fomento de la honestidad intelectual en sus estudiantes. Al igual que Peng (2012), Stodden (2015) y Gandrud (2015), en este artículo se consideró que una investigación es reproducible si cumple al menos con tres principios básicos, a saber, el método se describe de forma completa y detallada; los datos utilizados se ponen a disposición pública en medios adecuados y de acceso libre; el protocolo de depuración y tratamiento de los datos está documentado conforme a normas comúnmente aceptadas por comunidades epistémicas reconocidas.

## Método y materiales

Las características del problema, las interrogantes planteadas y el propósito del estudio demandaron el uso de un método de investigación cuantitativo. Se utilizó un diseño no experimental de carácter transversal debido a que no se manipularon variables de forma deliberada y a que solo interesó obtener, describir e interpretar datos en un momento único.

La población participante fue de 600 estudiantes universitarios que cursaban carreras de Ingeniería. Para su selección se usó un muestreo por conveniencia debido a la proximidad de los sujetos de estudio y a la escasa información relacionada con la utilización de métodos de reproducción en el ámbito educativo.

Se elaboró un cuestionario integrado por 14 ítems (Anexo 1) para el acopio de información. Los primeros siete se incluyeron con el propósito de obtener datos generales de los participantes; los cuatro siguientes se elaboraron para inspeccionar si contaban con antecedentes en tópicos estadísticos, metodológicos y de lenguajes de programación computacional; los ítems 11, 12 y 13 se confeccionaron para explorar en qué medida los participantes utilizaban los lenguajes de marcado LATEX y Markdown, así como el lenguaje de programación R para reproducir una investigación; el último ítem se elaboró para examinar en qué medida se promueve y utiliza la reproducibilidad de la investigación entre los estudiantes por parte de sus profesores.

Aún cuando las respuestas a los ítems no implicaron la realización de cálculo alguno ni la emisión de contestaciones explicativas, se consideró pertinente examinar su validez y consistencia en relación a la claridad de los términos utilizados. El cuestionario fue validado por el criterio de expertos y se pidió a 46 estudiantes de Ingeniería que participaran en una sesión presencial en que se dio respuesta a cada pregunta con el propósito de verificar la existencia o no de discrepancias en la interpretación para, de ser necesario, hacer los ajustes correspondientes.

Para sistematizar la información reunida, efectuar el tratamiento estadístico y elaborar este informe, se utilizaron las herramientas de cómputo siguientes:

- Entorno de Desarrollo Integrado (IDE) RStudio versión 1.0.136.
- Lenguaje de programación R versión 3.3.2 de 64 bits.
- Paquetes especializados en investigación reproducible adicionales al núcleo base de R:
  - knitr (1.15.1)
  - stargazer (5.2)
  - ggplot2 (2.2.1)
  - gridExtra (2.2.1)
  - cowplot (0.7.0)
  - ggrepel (0.6.5)
- Gestor de referencias Mendeley versión 1.17.6.
- Distribución Tex Live 2016 (LATEX) para sistema operativo macOS Sierra versión 10.12.2.
- Hoja de cálculo Microsoft Excel versión 15.27.
- Computadora MacBook Pro Mid 2014 con procesador Intel Core i5 de 2.6 GHz, 16 GB de memoria Ram y almacenamiento flash de 256 Gb.

El protocolo de trabajo se compuso de tres fases: [1] planeación y estructuración de un sistema de administración documental; [2] procedimiento de aplicación, acopio, depuración y sistematización de información; y [3] tratamiento estadístico de los datos. En la fase 1 se creó un sistema de administración para clasificar los archivos generados que después, a través de código de programación, permitió hacer reproducible este reporte de investigación. La estructura del sistema se construyó sobre la base de una carpeta principal en la que a su vez se crearon otras carpetas y archivos.

Data almacenó la base de datos con extensión `.csv`; Documents guardó documentos de carácter legal en el desarrollo de este estudio que no incidieron en el procedimiento investigativo, tratamiento e interpretación de resultados; Images integró las imágenes externas utilizadas en el reporte; Figures se generó automáticamente durante el proceso de compilación para almacenar las representaciones gráficas producidas. También se incorporó una base de datos con extensión `.bib` que incluyó las fuentes de información consultadas y un archivo con el código de programación de extensión `.Rnw`. En la Figura 1 se muestra la estructura general del sistema.

Para posibilitar la reproducción de los resultados obtenidos, la replicación del estudio en otros contextos y promover prácticas transparentes de investigación y programación computacional, se elaboró un archivo con la descripción global del informe.

En la fase 2 se seleccionaron seis instituciones universitarias, cinco de ellas especializadas en la formación de profesionales en el campo de la Ingeniería. Enseguida se aplicó el cuestionario a los participantes para que emitieran sus respuestas, se reunió la información y elaboró un archivo con extensión `.csv` que integró los registros.

En la fase 3 se utilizó el lenguaje de programación R para el tratamiento estadístico de los datos y la generación de representaciones gráficas. El código de programación usado se incrustó en trozos (*chunks*) dentro del documento `.Rnw`

## Resultados

La distribución por sexo fue 212 (35.33 %) y 388 (64.67 %) para femenino y masculino respectivamente. Estas cifras están acordes a la matrícula promedio de las instituciones participantes en las que más del 55% de los estudiantes inscritos en carreras de Ingeniería pertenecen al sexo masculino.

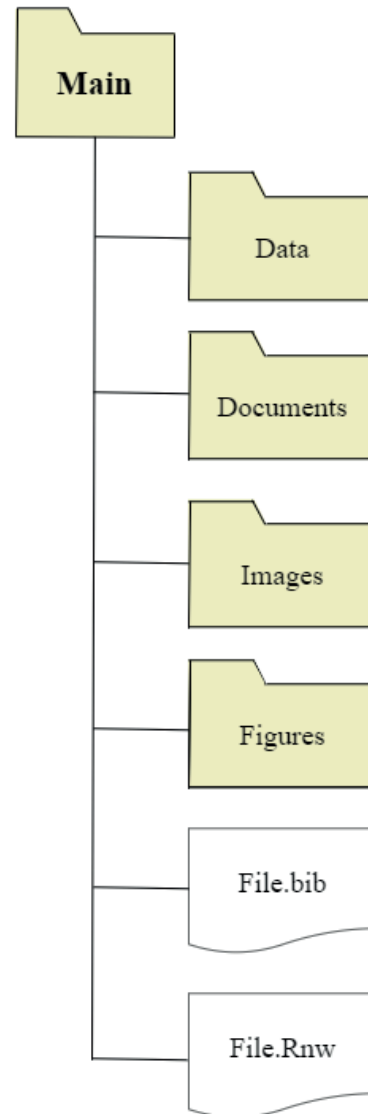


Figura 1: Estructura de archivos. Fuente: Elaboración propia.

En relación a la edad en años cumplidos, en términos globales el rango de valores fue nueve (17 a 26) y  $la = 20;330$  con una  $s = 1;881$ . En la Figura 2 se observa la dispersión y asimetría entre ambos sexos en esta variable. El promedio de edad fue más alto en el sexo femenino y la diferencia en sus medianas fue de un año aproximadamente. En ambos grupos se presentaron valores atípicos.

La mayoría de los estudiantes afirmó que cursaron materias relacionadas con métodos de investigación, estadística y lenguajes de programación computacional. La Tabla 1 presenta un resumen estadístico que destaca lo siguiente: [1] todos recibieron instrucción escolar en al menos un lenguaje de programación computacional; [2] al menos 350 cursaron dos o más materias relacionadas



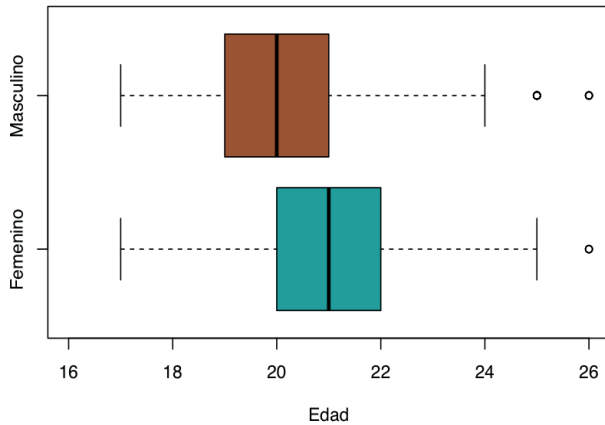


Figura 2: Edad de los participantes y su clasificación por sexo. Fuente: Elaboración propia.

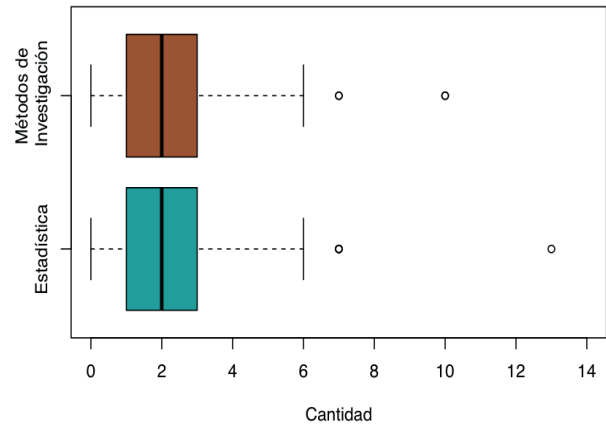


Figura 3: Comparación materias de Metodología y Estadística. Fuente: Elaboración propia.

con métodos de investigación y estadística; y [3] 150 o más afirmaron conocer el manejo de al menos un lenguaje de programación. (Tabla 1).

La Figura 3 muestra una comparación entre la cantidad de materias cursadas de estadística y métodos de investigación. Los valores de los cuartiles Q1, Q2 (mediana), Q3 y los límites de los bigotes son semejantes. Cuatro valores quedaron fuera del rango intercuartílico  $iqr = 2$ , dos de ellos (13 y 10) resultaron extremadamente atípicos porque excedieron el supuesto:  $x > (Q3 + 3(iqr))$  es decir, el valor es mayor que el resultado ubicado en el cuartil tres más el producto de tres veces el rango intercuartílico.

Es difícil saber las causas subyacentes de las desviaciones estándar  $s = 1;420$  y  $s = 1;697$  respectivamente; una hipótesis es que se deban a la formación académica recibida en el nivel bachillerato; otra hipótesis es la diferencia existente entre el número de materias en estas disciplinas que se incluyen las mallas curriculares de las carreras de Ingeniería que ya habían cursado los participantes al momento del estudio.

Respecto a la identificación, conocimiento y uso de herramientas computacionales necesarias para reproducir una investigación, los ítems 11, 12 y 13 permitieron saber si los participantes identificaban y usaban los lenguajes LATEX, Markdown y R. Las opciones de respuesta a dichos ítems fueron: [1] No sabía de su existencia (NSE); [2] Sabe que existe, pero no lo ha utilizado (SEPNU); [3] Está aprendiendo a usarlo (EAUS); [4] Lo ha usado antes, pero ya no lo hace (UAPNA); [5] Lo sigue usando (SU). En la Figura 4 se muestra un panorama general de los resultados más representativos, entre los que destacan:

LATEX: Aproximadamente 400 participantes no sabían de su existencia; menos de 100 lo identifican, pero nunca lo han usado; alrededor de 40 estaban aprendiéndolo y una cantidad muy baja que lo ha usado antes continúa haciéndolo.

Markdown: Más de 500 participantes desconoce su existencia; la cantidad de participantes que estaban aprendiéndolo fue muy baja y su uso fue prácticamente nulo entre quienes admitieron utilizarlo en el pasado. Estos resultados son importantes si se considera que es un lenguaje de marcado ligero muy publicitado y con una curva de aprendizaje corta.

Statistic	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
Edad	20.330	1.881	17	19	20	22	26
Lenguajes. cursados	1.133	0.340	1	1	1	1	2
Lenguajes que conoce	1.430	1.199	0	1	1	2	8
Métodos de investigación	2.032	1.420	0	1	2	3	10
Estadística	2.188	1.697	0	1	2	3	13

Tabla 1: Resumen estadístico para variables seleccionadas. Fuente: Elaboración propia.

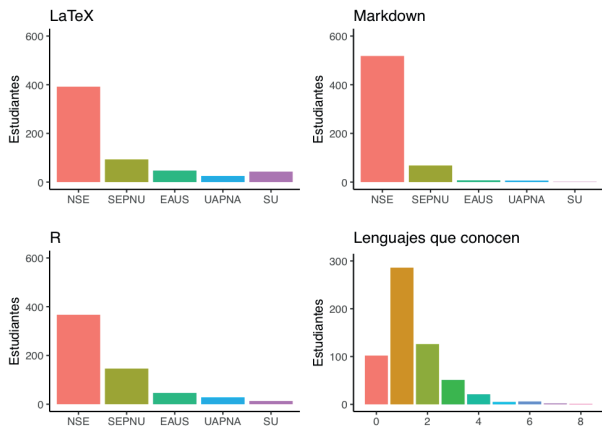


Figura 4: Lenguajes que identifican y usan. Fuente: Elaboración propia.

R: También es un lenguaje desconocido para una proporción alta de estudiantes (> 50%). Su utilización es muy baja y, al igual que LATEX, alrededor de 40 estaban aprendiendo a utilizarlo. (Figura 4).

El ítem 14 tuvo el propósito de examinar en que medida se promueve y usa la reproducibilidad de la investigación. Las opciones de respuesta fueron: [1] Entregue un reporte escrito únicamente (ER); [2] Entregue un reporte y una base de datos (ERBD); [3] Entregue un reporte, una base de datos y hoja con cálculos (ERBDC); [4] Entregue un reporte, una base de datos, hoja con cálculos y el código de programación desarrollado (ERBDCC); [5] Entregue una carpeta digital que contenga un reporte escrito o digital, la base de datos creada, imágenes, y el código de programación desarrollado que, al ejecutarse, permita reproducir los resultados (EDCOMP).

Se encontró que la exigencia de la reproducción completa es mínima en las aulas universitarias. Aún predomina la entrega de un reporte escrito como evidencia de la realización de un estudio, como se puede notar en la Figura 5.

## Conclusiones y recomendaciones

En la realización de este estudio se partió de las premisas que consideran la institución escolar y el profesorado como ejes centrales en la promoción de la reproducibilidad y fomento de la honestidad intelectual. En el caso de la reproducibilidad es deseable tener nociones básicas en algún lenguaje de programación; el grado de dominio en tópicos estadísticos y metodológicos estará en función del tipo de análisis y presentación que se exija en un estudio particular.

El lenguaje de marcado LATEX tiene una larga tradición –más de tres décadas– en el ámbito académico y

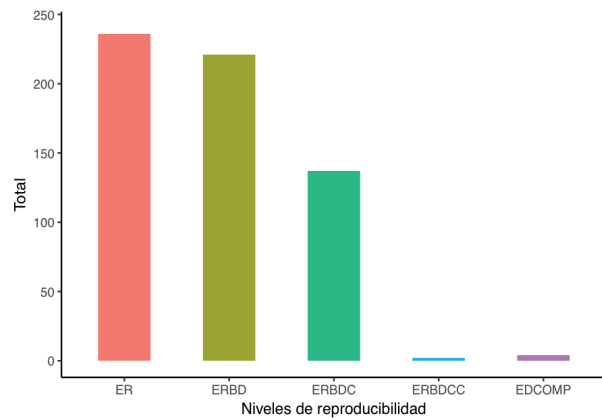


Figura 5: Promoción y uso de la reproducibilidad en la investigación. Fuente: Elaboración propia.

científico. Su uso es recomendado en la elaboración de informes que requieren tipografía especializada y portabilidad completa con independencia del sistema operativo. Markdown es un lenguaje de marcado ligero de reciente creación –marzo de 2004– que tiene como atributos principales su facilidad de uso e integración con lenguajes de programación de propósito general y cómputo científico. R es posiblemente la lingua franca de la estadística computacional, cuenta con la potencia de cálculo suficiente para satisfacer las exigencias de cualquier rama de la estadística y la elaboración de gráficos sofisticados. La combinación de los dos primeros con el tercero (LATEX–R) y (Markdown–R) constituyen una alternativa factible y viable de utilizarse en el ámbito universitario.

Actualmente existe una crisis de credibilidad en algunas disciplinas científicas debido a la imposibilidad de garantizar la veracidad e imparcialidad en la generación, sistematización e interpretación de datos observacionales y experimentales que se publican en medios especializados. Lo anterior ha incrementado la exigencia de que los reportes de investigación y sus resultados sean factibles reproducirse por evaluadores externos.

En el caso de este estudio, los resultados indican que la totalidad de los participantes han interactuado en el ámbito escolar con lenguajes de programación, tópicos estadísticos y métodos de investigación. Esta interacción no garantiza su articulación y uso efectivo en la solución de un problema o la atención de una necesidad en el ámbito ingenieril o social, pero al menos es un indicio de que existe una base teórica y procedimental en esos campos.

Respecto de los lenguajes, la mayoría de los participantes ignora su existencia y, salvo casos aislados, LATEX, Markdown y R no se enseñan en las carreras de Ingeniería.

ría. Al parecer la enseñanza se orienta a lenguajes de propósito general en detrimento de los orientados al cómputo científico.

Continúa el predominio de solicitar la entrega de un reporte escrito como evidencia única en la realización de un estudio cuantitativo; la entrega de bases de datos y código de programación tienen menor incidencia, y aún en los casos que se exige, es altamente probable que se haga de manera desarticulada, es decir, posiblemente se les solicite los archivos sin conexión alguna.

Se recomienda elaborar y someter a prueba modelos de trabajo para introducir la reproducibilidad como parte inherente de la formación profesional en el campo de la Ingeniería. Se sugiere que los modelos incluyan el planteamiento de proyectos que integren la triada: estadística-programación computacional-métodos de investigación desde una perspectiva holista en que se privilegie la solución de problemas y la atención de necesidades ingenieriles y sociales por igual.

## Referencias

- Baker, M. (2016). Is there a reproducibility crisis? *Nature*, (533), 452–454.
- Claerbout, J. y Karrenbach, M. (1992). Electronic Documents Give Reproducible Research a New Meaning. En *Meeting of the Society of Exploration Geophysics. Society of Exploration Geophysics*.
- Fomel, S. y Claerbout, J. (2009). *Reproducible Research. Computing in Science and Engineering*, 5–7.
- Gandrud, C. (2015). *Reproducible research with R and Rstudio* (2ª Ed.). USA: CRC Press.
- Ioannidis, J. (2005). Why most published research findings are false. *PLoS Medicine*, 2(8):696–701.
- Iqbal, S. (2016). Reproducible Research Practices and Transparency across the Biomedical Literature. *PLoS Biology*, 14(1),1–13.
- Open Science Collaboration (2015). Estimating the Reproducibility of Psychological Science. *Science*, 349(6251).
- Peng, R. (2012). Reproducible Research in computational Science. *Science*, 334(6060).
- Stodden, V. (2015). Reproducing Statistical Results. *Annual Review of Statistics and Its Application*, 2(1), 1–19.
- Xie, Y. (2015). *Dynamic Documents with R and knitr*. (2ª Ed.). USA: CRC Press.



## Anexo 1. Cuestionario

### Encuesta sobre el uso de herramientas computacionales en estudiantes universitarios

El propósito de este cuestionario es obtener información sobre herramientas computacionales por parte de estudiantes que cursan estudios profesionales en Instituciones de Educación Superior. Los resultados serán tratados con confidencialidad y para fines estadísticos.

**Instrucciones:** Responda los ítems que se enuncian a continuación marcando con una **X** la opción que corresponda.

**1. Sexo:**

- Femenino
- Masculino

**2. Edad:** \_\_\_\_\_

**3. ¿De cuál Sistema de Educación Media Superior egresó?**

- Cobach
- Dgeti
- Conalcp
- Dgeta
- Profeco
- Bachillerato privado
- Otra: \_\_\_\_\_

**4. ¿En qué institución cursa sus estudios profesionales?**

- Instituto Tecnológico de Chihuahua
- Instituto Tecnológico de Chihuahua II
- Instituto Tecnológico y de Estudios Superiores de Monterrey
- Universidad Autónoma de Chihuahua
- Universidad Politécnica de Chihuahua
- Universidad Tecnológica de Chihuahua

**5. Carrera estudia actualmente:**

- Ingeniería en Gestión Empresarial
- Ingeniería Industrial
- Ingeniería Sistemas Computacionales
- Ingeniería en Software
- Ingeniería en Ciencia Computacional
- Ingeniería en Informática
- Ingeniería en Matemáticas
- Ingeniería en Física
- Ingeniería Aeroespacial
- Ingeniería Química
- Química
- Química Bacteriólogo Parasitólogo
- Medicina
- Ingeniería Biomédica
- Ingeniería Mecánica Automotriz
- Ingeniería en Tecnología Ambiental
- Ingeniería en Aeronáutica
- Ingeniería en Procesos Industriales
- Ingeniería en Mecatrónica
- Ingeniería en Mantenimiento
- Ingeniería Mecánica
- Ingeniería en Energías Renovables
- Ingeniería Eléctrica
- Otra: \_\_\_\_\_

**6. Semestre o cuatrimestre que cursa: Por favor ante la modalidad:**

**7. ¿Le han enseñado algún Lenguaje de Programación en su trayectoria como estudiante?**

- Si
- No

**8. ¿Cuántos Lenguajes de Programación sabe utilizar al menos de forma elemental? \_\_\_\_\_**

**9. ¿Cuántas materias relacionadas con Métodos de Investigación ha cursado en su trayectoria como estudiante? \_\_\_\_\_**

**10. ¿Cuántas materias de Estadística ha cursado en su trayectoria como estudiante? \_\_\_\_\_**

**11. En relación al Lenguaje de Programación R, usted:**

- No sabía nada de su existencia.
- Sabe que existe, pero no lo ha utilizado.
- Está aprendiendo a usarlo.
- Lo he usado antes, pero ya no lo hago.
- Lo sigue usando

**12. Respecto del Sistema de Composición de Textos, usted:**

- No sabía nada de su existencia.
- Sabe que existe, pero no lo ha utilizado.
- Está aprendiendo a usarlo.
- Lo he usado antes, pero ya no lo hago.
- Lo sigue usando

**13. En relación al Lenguaje de Marcado Ligero Markdown, usted:**

- No sabía nada de su existencia.
- Sabe que existe, pero no lo ha utilizado.
- Está aprendiendo a usarlo.
- Lo he usado antes, pero ya no lo hago.
- Lo sigue usando

**14. Cuando se le pide que realice un estudio que implica el uso de cálculos matemáticos y la entrega de un informe con los resultados, su profesor (a) le solicita:**

- Entregue un reporte por escrito únicamente.
- Entregue un reporte y una base de datos.
- Entregue un reporte, una base de datos y hoja con cálculos.
- Entregue un reporte, una base de datos, hoja con cálculos y el código de programación desarrollado.
- Entregue una carpeta que contenga un reporte escrito o digital, la base de datos creada, imágenes y el código de programación desarrollado que, al ejecutarse, permita reproducir los resultados.

**Responsable del estudio:**

Dr. Jesús Humberto Cuevas Acosta  
Departamento de Ciencias Básicas  
Instituto Tecnológico de Chihuahua II